Symposium

Journal of INTERNAL MEDICINE

doi: 10.1111/j.1365-2796.2011.02493.x

Systems medicine and metabolic modelling

A. Mardinoglu & J. Nielsen

Department of Chemical and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden[1]

Abstract. Mardinoglu A, Nielsen J (University of Technology, Gothenburg, Sweden). Systems medicine and metabolic modelling (Key Symposium). *J Intern Med* 2012; **271**: 142–154.

Several complex diseases are caused by the malfunction of human metabolism, and deciphering the underlying molecular mechanisms can elucidate their aetiology. Systems biology is an integrative approach combining experimental and computational biology to identify and describe the molecular mechanisms of complex biological systems. Systems medicine has the potential to elucidate the onset and progression of complex metabolic diseases through the use of computational approaches. Advances in biotechnology have resulted in the provision of high-throughput data, which provide information about different metabolic processes. The systems medicine approach can utilize such data to reconstruct genome-scale metabolic models that can be used to study the function of specific enzymes and pathways in the context of the complete metabolic network. In this review, we outline the importance of genome-scale models in systems medicine and discuss how they may contribute towards the development of personalized medicine.

Keywords: genome-scale metabolic models, metabolism, personalized medicine, systems biology, systems medicine.

Introduction

For many complex human diseases, such as cancer, cardiovascular and neurodegenerative diseases, it remains a challenge to find effective treatment strategies because of the complexity of the biological networks that underlie onset and progression of the disease. For such complex conditions, there are often many different mechanisms that result in the same phenotype, and it is therefore difficult to identify the exact cause of the disease and design adequate and efficient treatment strategies. Considering the complexity of both the cellular networks that are perturbed in connection with disease development and progression and the complex interactions between different cell types in the human body, it is generally difficult to dissect the molecular causes of different phenotypic developments. Even though large data sets are accumulating, such as omics data (e.g. genomics, transcriptomics, proteomics, metabolomics, fluxomics and bibliomics data) from different tissues and in large patient cohorts, it is not straightforward to use these data to identify the root causes of disease development. However, in recent years, there has been progress on the use of computer models to integrate large data sets as well as for simulation of biological networks with the objective of studying their dynamic behaviour [1–3]. This approach is referred to as systems biology [4].

Systems biology is an integrative research strategy that combines experimental and computational biology to identify the molecular mechanisms underlying the components of a complex biological system and to obtain a quantitative description [1, 5, 6]. This rapidly developing field integrates mathematical models with publicly available experimentally observed high-throughput data, such as genomes, transcriptomes, proteomes and metabolomes, and reconstructs biochemical networks for systematic analysis of complex systems using interdisciplinary computational tools [7]. In the context of systems biology, the computational approach focuses on the dynamics and interplay between biological systems such as cells, tissues and organs using a holistic approach rather than reductionism that focuses on individual components and typically excludes information regarding time, space and context [8].

In systems biology, predictive mathematical models are employed for the analysis of given experimental data in a quantitative fashion, for gaining new biological knowledge or for performing predictive simulations. There are two approaches to systems biology: the top-down and the bottom-up approach [5]. The top-down approach is a data-driven process where high-throughput experimental data are analysed with the objective of finding patterns or the function of biological subsystems (or pathways) in

the system being studied. By contrast, the bottom-up approach is typically hypothesis driven where detailed knowledge of subsystems is reconstructed into a mathematical model that can be used to describe the whole system. The advantage of the top-down approach is that it does not rely on prior knowledge of the system or subsystem; however, owing to the high dimensionality of biological systems, correlation analysis may not be necessary to identify biologically relevant causal relationships between variables. Nevertheless, if high-throughput data are combined with integrated network reconstructions, it is possible to identify biologically meaningful correlations between the many components in a biological system [9], and hence there is much interest in establishing reconstructed biological networks that can drive data analysis. Because of interest in the reconstruction of cellular pathways into *in silico* representation and application of such networks to biological systems, standardized programming languages and tools have been developed and knowledge of graph theory has been employed for their representation [10, 11].

In order to extend the recent developments in systems biology to medical applications, it is necessary to reconstruct and analyse large-scale biochemical, metabolic, signalling, protein, microRNA and gene regulatory networks that control human cellular processes [12]. The reconstruction of such biological network models, the combination of these models with omics data and their application to specific medical questions are often referred to as systems medicine. It is expected that this approach will revolutionize medical research in the near future [13, 14]. The development of systems medicine is enabled by the continuous progress of high-throughput experimental and computational technologies in medical genomics and bioinformatics [8, 12]. Systems medicine allows for a better understanding of the structure and function of the human genome and its associations by determining the links between genotypes, phenotypes and environmental factors (e.g. diet and exposure to toxins) [15] (Fig. 1). The integrative systems medicine research approach helps to understand the behaviour of the human body at all levels of organization by analysing its different constituents. Subsequently, it offers the prospects of modelling complex diseases, establishing novel diagnostic and therapeutic techniques [16], identifying new drug targets [17], developing a system-orientated drug design strategy [18] and eventually achieving effective personalized medicine [19, 20].

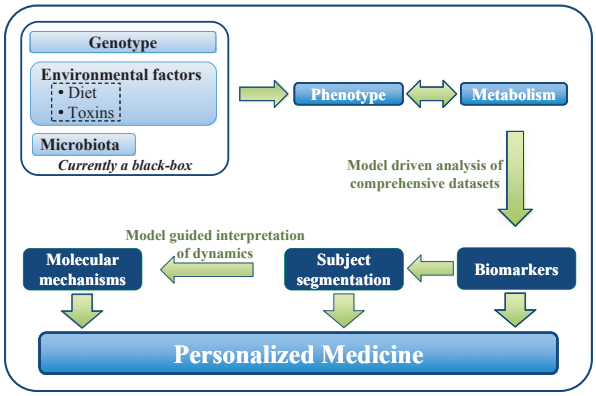

Fig. 1 *Schematic diagram to illustrate how systems medicine may contribute to the development towards personalized medicine. Phenotype is determined by genotype, environmental factors (such as dietary intake and exposure to toxins) and interactions with the microbiome. The phenotype is reflected by metabolism, which involves a number of processes including digestion of the diet, degradation of toxins and metabolic interaction between the gut microbiome. Based on model-driven analysis of comprehensive data sets, it may be possible to dissect the metabolic functions in different subjects and thereby identify biomarkers associated with specific phenotypes. This may allow for subject segmentation, which is essential for improved clinical trials as well as improved treatment through personalized medicine. The biomarkers will continue to play a role in diagnosis. Based on detailed patient segmentation, it will also be possible to obtain homogeneous subject groups, which can be used for more detailed clinical studies with the objective of performing model-guided interpretation of dynamic responses to different treatment strategies, and this may lead to the identification of molecular mechanisms underlying different diseases. This can be useful for drug discovery as well as improved disease treatment.*

The systems medicine approach also allows the establishment of new associations between biological functions and a wide range of diseases (e.g. immunological, inflammatory, infectious, neurological and respiratory diseases), including human metabolism-related conditions (e.g. obesity, diabetes, hypertension, hyperinsulinaemia, dyslipidaemia and certain types of cancer) [12]. A better understanding of human metabolism could explain the onset and progression of such diseases, as they are most commonly caused by irregular functioning of energy metabolism in different tissues and/or cell types in the human body. However, metabolism is highly complex and involves a very large number of metabolic reactions in different tissues. The control and regulation of human metabolic processes could be explained through the use of computational modelling and the recon-

struction of genome-scale metabolic models (GEMs). GEMs have a promising role in the analysis of the specific state of human tissues and can be employed as scaffolds for the development of personalized medicine, in line with the systems medicine approach.

In this review, we focus on the modelling of human metabolism, in the context of systems medicine, with the objective of gaining a better understanding of molecular mechanisms underlying different phenotypes. We will discuss the significance of personalized medicine for future health care, give a brief account of common metabolic disorders, outline the publicly available databases for reconstruction of GEMs and describe existing GEMs. Finally, we will consider how novel model-building algorithms may be applied for automated reconstruction of tissue-specific human metabolic network models.

Personalized medicine for future health care

Personalized medicine is a novel approach to health care that may enable customized patient-specific strategies based on each person's unique profile in terms of genetic, phenotypic and environmental factors [20]. The aim of this emerging field is not only to design patient-specific drug therapy based on the differences amongst individuals, but also to personalize treatment in order to account for patient differences in disease onset, course and development and drug response [21].

In order to achieve personalized medicine in an effective way, healthcare providers must translate their predictions on the basis of genomic, clinical and environmental information into precise diagnostic tests that can then be used to define targeted therapies. Thereby personalized medicine allows patients and clinicians to focus on prevention and prediction of diseases, to make informed medical decisions and to provide a higher probability of preferred outcomes, a reduced probability of side effects and early detection of disease interference. This will move costs from treatment to diagnosis, but is still expected to result in significantly reduced healthcare costs. In cancer patients, personalized medicine has already been partly successful by determining the probability of having a serious adverse reaction to certain cancer drugs. Nevertheless, results are still difficult to fully translate into clinical practice. Furthermore, for some genetic diseases, such as muscular dystrophy, cystic fibrosis and sickle cell anaemia, the associated influencing individual genes have already been found based on genetic research conducted 50 years ago. However, these single gene-related disorders are not typical and can be influenced by other genes, microbiota and environmental factors as well as by their interplay.

A key requirement for the introduction of personalized medicine to the future healthcare system is the identification, classification, validation and clinical use of new diagnostic, prognostic and predictive biomarkers. Systems medicine provides the possibility of development of new specific biomarkers by paying particular attention to the biomarker source. Although large amounts of omics data are collected from different tissues in health and disease states, simple statistical-based correlation analysis has not yet enabled the extraction of specific biomarkers associated with different diseases [22]. Reconstruction of biological networks (e.g. GEMs) can lead to identification of altered molecular processes in disease development, and this can drive the search for novel biomarkers (Fig. 1). Particular metabolic pathways showing up-regulation at the transcriptional level could point to pathway intermediate metabolites expected to be present at elevated levels, perhaps even in the plasma. Clearly, this kind of analysis will require extensive data obtained from analysis of biopsies from different subjects, but it may only be necessary to perform this analysis using a few carefully selected subjects, followed by validation analysis of the proposed biomarkers in a larger cohort. We therefore predict that metabolic networks will be instrumental in driving the identification of metabolite-based biomarkers.

Metabolism-related complex diseases

Human metabolism-related diseases have been recognized as an emerging, global public health problem. However, it is difficult to identify the exact mechanisms underlying the onset and progression of metabolism-related disorders because they are often complex, involve interplay between different tissues, have strong multigene components and have many different underlying mechanisms that result in the same gross phenotype. The most common human metabolism-related disorder is obesity, which may lead to many different types of complications and plays a critical role in the pathogenesis of a multitude of diseases. The occurrence of obesity continues to increase dramatically, and it is becoming one of the most common causes of death worldwide [23]. The prevalence of obesity is increasing in adults and children in developed countries, and it is expected to be one of the most serious public health problems of the

21st century. Obesity is most commonly caused by the combination of genetic predisposition, epigenetic regulation, unhealthy environmental factors and individual behaviours (including extreme food and energy intake and lack of physical activity), and their interactions [24]. This complex condition is also associated with insulin resistance and with the development of the metabolic syndrome, which is a combination of medical disorders that increases the risk of developing cardiovascular disease and type 2 diabetes. Insulin resistance is often seen in people with visceral adiposity, and there may be a link to the metabolic syndrome [25, 26]. First, visceral adipose cells produce significant amounts of pro-inflammatory cytokines that disrupt the normal action of insulin in fat and muscle cells. Second, visceral adiposity is related to an accumulation of fat in the liver, a condition known as nonalcoholic fatty liver disease (NAFLD), and this results in excessive release of free fatty acids into the bloodstream and an increase in hepatic glucose production, which are both associated with the metabolic syndrome. In this context, it is also interesting to look at the role of diet in NAFLD disease progression as research suggests that increased intake of long-chain n-3 polyunsaturated fatty acids (PUFAs) decreases lipid accumulation, inflammation and injury to hepatocytes. Conversely, NAFLD has been associated with reduced levels of hepatic long-chain n-3 PUFAs. Obesity also has adverse effects on health that could lead to reduced life expectancy and is associated with various common diseases, particularly obstructive sleep apnoea, certain types of cancer and osteoarthritis [27].

For a better understanding of how the metabolic syndrome, insulin resistance, type 2 diabetes, cardiovascular disease and dyslipidaemia are related, it is necessary to include the many different possibilities into mathematical models that can then be used for the evaluation of the different putative mechanisms. Several investigators have focused on the glucose–insulin control system and developed phenomenological models such as maximal and minimal models [28]. Maximal models attempt to implement the body of knowledge about metabolic regulation with a large number of parameters using experimental data; by contrast, minimal models describe the key components of the system functionality and cannot estimate the values of all system parameters *in vivo* based on dynamic data. In minimal models, each of the parameters is tested and numerically identified from experimental data for a better explanation of the changes in the body caused by diabetes. In addition, insulin action is modelled by glucose kinetics at steady state by using tracer theory and linear time-dependent data, and at non-steady state by combining multiple steady-state tracer studies. It should be noted that both models have to consider not only the insulin and glucose systems but also simultaneously how they interact. So far, there have not been any applications of GEMs for delineating how metabolism in different tissues contributes to the development of metabolic syndrome and related diseases, but clearly detailed metabolic analysis of different tissues in subjects with different gross phenotypes may contribute to improved understanding of the underlying molecular mechanisms and further point to the identification of novel biomarkers that can be used for improved patient stratification (Fig. 1). In this way, it will be possible to design improved intervention strategies that are tailored to individual patients, rather than the current approaches in which all subjects with a certain phenotype (e.g. insulin resistance) are treated using the same strategy.

Reconstruction of genome-scale metabolic models

As metabolic diseases are a consequence of the failure of human metabolic processes, there has naturally been a focus on human metabolism and its regulation to determine the molecular mechanisms of these diseases. Metabolism is complex, involving a huge number of reactions, and reconstruction of human metabolic network models using a systems biology approach may therefore provide an excellent scaffold for this type of analysis. One of the key components of this approach is the construction of GEMs using a bottom-up reconstruction strategy in which all the available information about human metabolism is integrated [1, 9, 29, 30]. GEMs are very suitable for obtaining insights into metabolic phenotypes that emerge from complex biological systems and they further allow the integration of different data sets in a single modelling structure [31–33]. Such models represent a collection of metabolic reactions and associated enzymes in a specific cell, tissue or organism and enable the role of individual reactions and/or pathways to be analysed in the context of the entire network. GEMs are normally described mathematically by a stoichiometric matrix, S, that captures the stoichiometry of the reactions (Fig. 2) and provides quantitative information on how the different metabolites are linked to each reaction in the network [34]. In the S matrix, every reaction is represented by a column, and each row represents a metabolite. In order to generate functional descriptions of network properties, matrix analysis methods are used after defining the network boundaries [35], and these stoichiometric models are employed for simulation of

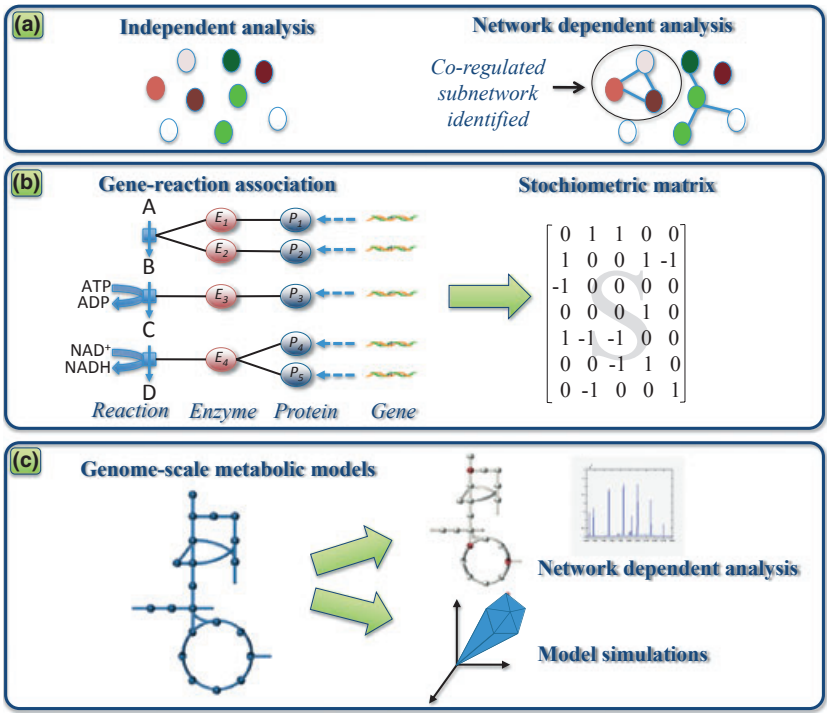

Fig. 2 *Schematic diagram to illustrate the process of reconstruction of GEMs and their use for simulation and analysis. (a), Biological information about how the different components are connected is not used in traditional statistical analysis of high-throughput data. Metabolic networks represent highly annotated interaction graphs that can be used for improved statistical analysis of high-throughput data, for example to identify marker metabolic pathways and/or functions in terms of so-called subnetworks. (b) GEMs provide gene–protein–enzyme reaction links. Furthermore, they provide detailed stoichiometric information on individual enzymatic reactions. This stoichiometric information is represented mathematically in a stoichiometric matrix, S, that includes information about how metabolites are linked to each reaction and provide linear relationships between the flux rates of the reactions and the metabolites. In order to simulate the functional properties of a tissue type using the S matrix, it is necessary to use tools from linear algebra. (c) The S matrix and information related to the reactions including gene, protein and enzyme associations are combined for reconstruction of comprehensive GEMs by employing a machine-readable format, the so-called Systems Biology Markup Language (SBML), for their representation. These reconstructed models can be used for simulation of metabolic functions using flux balance analysis as well as for improved statistical analysis for the identification of marker pathways or functions.*

growth and product formation by the use of computational analysis techniques such as flux balance analysis. The reconstruction process for GEMs has been reviewed extensively elsewhere [36, 37], and the total number of reconstructed GEMs for biological systems rises each year [38]. To date, GEMs have primarily been used for modelling microorganisms, for modelling the behaviour of cells grown in a laboratory environment and for integrative data analyses [34, 39], but recently they have also been used for analysis of data from clinical studies and from experimental studies in mice [40].

As GEMs provide gene–protein reaction links, they can also be used to bridge the gap between experimental biomedical research and patient-oriented research because cell type–specific GEMs can be used as scaffolds for data analysis (Fig. 3). These integrated models provide information by combining all biological, genomics and medical information on genes, gene-driven reactions, related diseases and drug targets at a systems level, and thereby allow the study of relationships between networks, functions, diseases and patients in the context of omics data. These models can be directly employed for understanding how the transcriptome and/or proteome influences different metabolic functions within the entire metabolic network, and thereby understand the metabolic responses to different disease states caused by disorders of tissue functions. Hence, these

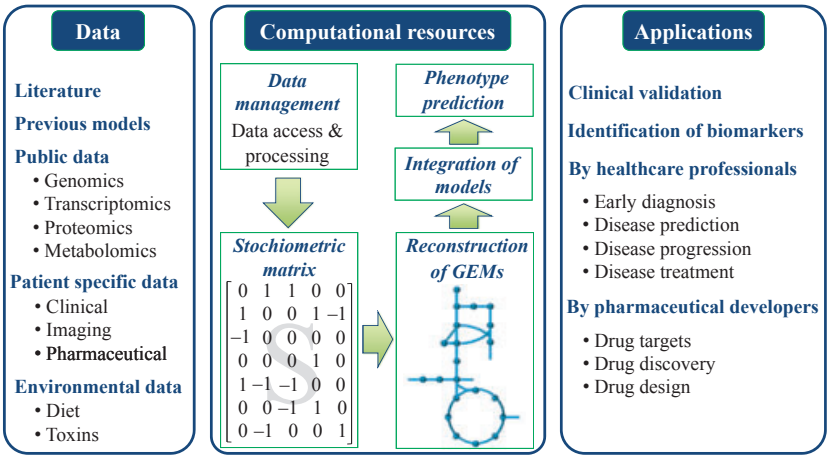

Fig. 3 *Schematic diagram to illustrate how GEMs may impact systems medicine. GEMs are reconstructed by incorporating information from published literature, patient-specific and environmental data and publicly available databases, resulting in tissue-specific models for all functions in the human body. Published literature contains experimental knowledge of the biochemical and physiological details of pathways and reactions, whereas public databases contain information about high-throughput data derived from genome annotation. These complex and varying data types are integrated into a single network for phenotype prediction using GEMs. Reconstructed GEMs can be validated mathematically by simulating known tissue functions, and thereafter clinical usage can be attained by testing the predicted phenotype. Following computational analysis and validation, the models can be used by healthcare professionals to understand the onset, progression and treatment of diseases, and by pharmaceutical developers to identify new drug targets and generate novel concepts for drug discovery and design. Furthermore, GEMs can be employed for the identification of biomarkers to achieve effective personalized medicine.*

integrated models may help clinicians to improve diagnosis and to generate the knowledge required to identify new therapeutic strategies (Fig. 3). Furthermore, these models may be of assistance within the pharmaceutical industry on several fronts by identifying novel drug targets, improving insight into the modes of action of existing drugs, enabling unsuccessful drug development projects to be terminated before they become too costly, facilitating early diagnosis in order to shift from disease treatment to disease prevention, and enabling personalized treatment, thus avoiding the prescription of drugs to nonresponders or to patients likely to suffer severe side effects.

Many different types of information can be used to reconstruct GEMs for human cell types, and these include previously published human-related GEMs and genomic and molecular information available in public databases such as Kyoto Encyclopedia of Genes and Genomes (KEGG) [41], HumanCyc [42], Reactome [43], BioCarta [44] and Rhea [45], as well as DNA microarray data for different tissues (collected from gene expression omnibus [46] and ArrayExpress [47]), the Human Protein Atlas (HPA) [48] and the Human Metabolome Database (HMDB) [49] (Table 1). Through combining information from these and other clinical databases and manually evaluating original research papers, high-quality, simulation-ready, tissue-specific GEMs for human tissue and cells can be generated by ensuring standardization of the data from the different sources such as Ensembl [50], Entrez [51], UniProt [52] and KEGG [41] for genes, proteins and compounds.

Kyoto Encyclopedia of Genes and Genomes is one of the most commonly used databases in connection with reconstruction of GEMs. It is a collection of well-known online databases and consists of three different types of data: genomic, chemical and network information. The data are linked with each other and with other existing publicly available databases (Table 1). The KEGG database has been developed as a computer representation of biological systems [53], and it can be employed for modelling, browsing and retrieval of data as a part of systems biology studies. KEGG integrates information on molecular interaction networks (e.g. pathways and complexes through pathway databases), information about genes and proteins generated by genome projects through gene databases, and information about biochemical compounds and reactions through compound and reac-

Table 1 *Publicly available resources for reconstruction of human genome-scale metabolic networks*

Resource	Web address	Description	Reference(s)
Metabolic pathways and reactions			
KEGG	http://www.genome.jp/kegg	Kyoto Encyclopedia of Genes and Genomes (KEGG) enables computational prediction of higher-level complexity of cellular processes and organism behaviours from genomic and molecular information.	[41]
HumanCyc	http://biocyc.org	HumanCyc: Encyclopedia of Homo sapiens Genes and Metabolism is a bioinformatics database that describes human metabolic pathways and the human genome.	[42]
Reactome	http://www.reactome.org	Reactome is an online encyclopedia of core human pathways including DNA replication, transcription, translation, cell cycle, metabolism and signalling.	[43]
BioCarta	http://www.biocarta.com	BioCarta simplifies the process of developing reagents to study new proteins, resulting from the explosion in the growth of new targets for academic study and drug development.	[44]
Rhea	http://www.ebi.ac.uk/rhea	Rhea is a freely available, manually annotated database of chemical reactions.	[45]
Genes and proteins			
HPA	http://www.proteinatlas.org	The Swedish Human Protein Atlas (HPA) project has been set up to allow for a systematic exploration of the human proteome using antibody-based proteomics.	[48]
GEO	http://www.ncbi.nlm.nih.gov/geo	The Gene Expression Omnibus (GEO) is a database that contains high-throughput gene expression data and hybridization arrays, chips and microarrays.	[46]
ArrayExpress	http://www.ebi.ac.uk/arrayexpress	The ArrayExpress Archive is a database of functional genomics experiments including gene expression.	[47]
Protein database	http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein	The protein database is a collection of sequences from several sources including translations from annotated coding regions in GenBank [75], RefSeq [76] and third-party annotation [77] sequence databases.	[78]
Entrez	http://www.ncbi.nlm.nih.gov/sites/entrez?db=gene	Entrez maintains information about genes from genomes of interest to the RefSeq [76] group.	[51]

Table 1 *(Continued)*

Resource	Web address	Description	Reference(s)
Ensembl	http://www.ensembl.org/index.html	Ensembl data set consists of gene models built from the genewise alignments of the human proteome as well as from alignments of human cDNAs using the Exonerate cdna2genome model.	[50]
Uniprot	http://www.uniprot.org	Uniprot is a comprehensive and high-quality resource for protein sequence and functional information.	[52]
wwPDB	http://remediation.wwpdb.org	The Worldwide Protein Data Bank (wwPDB) maintains a single Protein Data Bank Archive of macromolecular structural data.	[79]
ExPASy proteomics server	http://www.expasy.org/proteomics	The Expert Protein Analysis System (ExPASy) proteomics server is the new Swiss Institute of Bioinformatics (SIB) portal that is dedicated to the analysis of protein sequences and structures. It provides access to scientific databases and software tools in proteomics.	[80, 81]
HGNC	http://www.genenames.org	The HUGO Gene Nomenclature Committee (HGNC) is the only worldwide authority that assigns standardized nomenclature to human genes.	[82]
OMIM	http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim	Online Mendelian Inheritance in Man (OMIM) is a comprehensive, authoritative and timely compendium of human genes and genetic phenotypes.	[83]
Enzymes			
IUBMB	http://www.chem.qmul.ac.uk/iubmb/enzyme/	Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) database contains general information on enzyme nomenclature.	[84]
ExplorEnz	http://www.enzyme-database.org	ExplorEnz – The Enzyme Database is developed as a new way to access the data of the IUBMB Enzyme Nomenclature List.	[85]
BRENDA	http://www.brenda-enzymes.org	BRaunschweig ENzyme Database (BRENDA) is a comprehensive enzyme information system and the main collection of enzyme functional data available to the scientific community.	[86]
ExPASy-ENZYME	http://au.expasy.org/enzyme	ENZYME is a repository of information relative to the nomenclature of enzymes and based on the recommendations of the Nomenclature Committee of the IUBMB.	[80, 81]
Metabolites and compounds			
HMDB	http://www.hmdb.ca	The Human Metabolome Database (HMDB) contains detailed information about small-molecule metabolites found in tissues in the human body.	[49]

Table 1 *(Continued)*

Resource	Web address	Description	Reference(s)
PubChem	http://www.ncbi.nlm.nih.gov/sites/entrez?db=pccompound	The PubChem compounds database contains validated chemical depiction information provided to describe substances in PubChem substance.	[87]
ChEBI	http://www.ebi.ac.uk/chebi	Chemical Entities of Biological Interest (ChEBI) is an ontological classification of molecular entities focused on small chemical compounds.	[88]
PDB-CCD	http://remediation.wwpdb.org/ccd.html	The Protein Data Bank Chemical Component Dictionary (PDB-CCD) is an external reference file describing all residue and small-molecule components found in PDB entries.	[79]
LIPID MAPS	http://www.lipidmaps.org	Lipidomics Gateway is a comprehensive database for researchers interested in lipid biology. It is used to explore the rich information collections, tools and resources from the LIPID MAPS consortium.	[89]

tion databases. The KEGG pathway database also records information related to metabolism, genetic information processing, environmental information processing, cellular processes, human diseases and drug development [54].

Moreover, several other databases such as the HPA and HMDB can be used for general educational applications as well as within metabolomics, clinical chemistry and biomarker discovery by linking chemical, clinical and molecular biology/biochemistry data. The HPA database includes a systematic exploration of the human proteome using antibody-based proteomics by combining high-throughput generation of affinity-purified antibodies with protein profiling in a multitude of tissues and cells assembled in tissue microarrays [48, 55–57]. The HPA database is extremely rich in terms of protein location in different human cell types and cell lines, and it has recently been expanded to cover the intracellular localization of proteins [58]. The HMDB database contains detailed information about small-molecule metabolites found in various human cell types [49] and hence provides evidence for the presence of metabolic functions in the different cell types.

Published genome-scale metabolic models

GEMs represent a comprehensive collection of metabolic information, that is, which metabolites are present in a given cell, how they are inter-converted, which enzymes catalyse these inter-conversions, and which genes encode for the enzymes. In the reconstruction of GEMs, published information about specific metabolic functions is often used, and hence these models also represent structured databases of published literature about specific metabolic functions. The first attempts to generate GEMs for human cells were made in 2007. Two generic stoichiometric networks of human metabolism such as Recon 1 [59] and the Edinburgh human metabolic network (EHMN) [60] were manually reconstructed using over 50 years of legacy data. Recon 1 and EHMN are the first global biochemical human metabolic networks that can be employed for systematic studies of metabolism and identification of gaps in our understanding of metabolism. Recon 1 includes the compartmentalization of metabolites in seven intracellular locations, as well as their exchange between these compartments, whereas EHMN first included metabolite compartmentalization in its second version in 2010 [61]. Both models use confidence scores and literature references based on known biological evidence associated with each gene, protein and reaction by using

precise Boolean descriptions for gene–protein relationships. Both networks also enable the computational analysis of properties of the entire human metabolic network and provide frameworks for analysis and interpretation of high-throughput data. Furthermore, these networks provide the possibility of illustrating the underlying mechanisms of metabolism-related diseases.

Even though these first reconstructions represent a major advance, human metabolism is of course very specialized in different tissues and cell types, and there is therefore a need to generate cell type–specific GEMs, which require much more effort than the reconstruction of the first generic human models [59–61]. Shlomi *et al.* [62] overcame this problem by considering gene expression measurements and generated draft reconstructions of tissue-specific metabolic networks in humans by mapping tissue-specific expression-profiling data against the reconstruction of the global human metabolic network. In this way, they generated draft GEMs for 10 different tissues, but still these GEMs are not specific for the many different cell types that are present in some tissues. In another attempt to generate specific GEMs, proteomic data from human cardiac mitochondria were used to form a tissue-specific organelle model [63].

Recently, two research groups independently published tissue-specific GEMs for hepatocytes [64, 65]. These models represent important manual reconstructions that will allow for benchmarking of other approaches; for example, where high-throughput data are used to semi-automatically reconstruct tissue- or cell type–specific GEMs. In addition to the development in reconstruction of tissue-specific models, Lewis *et al.* [66] presented a work flow for modelling human metabolism within and between different types of cells, and this workflow was applied to modelling energy metabolism in the brain considering the metabolic interactions between astrocytes and various neuron types related to Alzheimer's disease. In addition, a cell-specific alveolar macrophage model that was used for simulation of metabolic exchanges between host (the alveolar macrophage) and pathogen (*Mycobacterium tuberculosis*) during infections was reconstructed [67].

Besides the use of GEMs for analysis of human cell types, these models can also be used for simulation of how the many different types of bacteria in the gut microbiome contribute to overall gut function [68, 69]. The gut microbiome constitutes a very complex biological system that is difficult to analyse but, through the reconstruction of GEMs for key bacteria in this complex ecological system, it will be possible to simulate and evaluate different hypotheses concerning the function of the individual bacteria in the community.

Future developments

Even though there have been some developments towards the reconstruction of tissue- or cell type–specific GEMs (Table 2), there is still an enormous need to improve the existing models and generate new models for all key human tissues and/or cell types. It should also be noted that even the most complete GEMs are not perfect and may contain gaps or missing information [70], which can result from missing information about reactions between metabolites. To generate fully connected networks, a gap-filling process may have to be used [71], and recently different computational tools have been presented for this process [72–74]. However, further developments are needed in order to allow for automatic (or semi-automatic) reconstruction of tissue- or cell type–specific GEMs.

Our group has recently embarked on the so-called Human Metabolic Atlas project that will allow for automatic reconstruction of 69 preliminary cell type–specific GEMs from generic human models by

Table 2 *Human genome-scale metabolic models*

Model	Body site	Compartments	Genes	Metabolites	Reactions	Reference
Recon 1	Global	8	1496	2712	3843	[59]
EHMN (2007)	Global	NA	2322	2671	2823	[60]
EHMN (2010)	Global	8	2699	2634	6216	[61]
HepatoNet1	Hepatocytes	9	NR	777	2539	[64]
Computational liver model	Hepatocytes	8	NR	1360	1827	[65]
IAB-AMQ-1410	Alveolar macrophage	8	1410	2583	3394	[67]

NA, not applicable; NR, not reported; EHMN, Edinburgh human metabolic network.

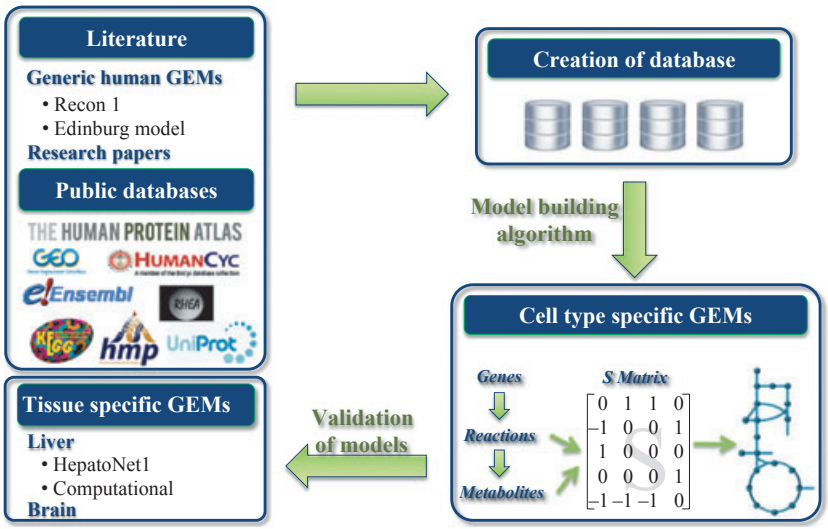

Fig. 4 *Schematic diagram to illustrate how we propose to perform automated reconstruction of cell type–specific GEMs, which are likely to play an important role in the future development of systems medicine. After representation of global human metabolic networks, several tissue-specific GEMs for different tissues, such as liver and brain, are reconstructed. However, existing GEMs do not fully represent the molecular mechanisms of related tissues. Therefore, there is still an enormous need to improve the existing tissue-specific GEMs and reconstruct draft GEMs for other tissues. In this context, a gene reaction database is generated by incorporating the existing human metabolic networks and by employing the public databases for standardization of the data. Next, a model-building algorithm is applied to the gene reaction database to reconstruct connected cell type–specific draft models. Knowledge of the existence of proteins in each cell type in the Human Protein Atlas (HPA) and of metabolites in the Human Metabolome Database (HMDB) is used during the reconstruction. For validation of the model-building algorithm, generated cell type–specific models can be used for the simulation of known functions of related cell type and compared with published corresponding tissue-specific models.*

using publicly available transcriptome, metabolome and tissue-specific protein localization data (Fig. 4). Key data for this process are from the HPA database [48], which provides information about the presence of proteins in 69 different cell types in different human tissues; however, we are also using tissue-specific microarray [46, 47] and metabolome data [49]. The automatic reconstruction process is based on generic human models such as Recon 1 [59] and EHMN [61], but reaction information from KEGG [41] and HumanCyc [42] is also parsed into a newly generated in-house database, in which all information is available in a standardized way by collecting data from Ensembl [50], UniProt [52] and KEGG [41]. A model-building algorithm is applied for integrating all this information and generating the cell type–specific models. The manually reconstructed GEMs for hepatocytes will be very valuable for validation of the model-building algorithm, as these are reconstructed based on the review of more than 1500 original research publications [64]. The reconstructed GEMs will be made publically available on a website and will eventually allow for integrative analysis and hence may become a valuable tool for analysis and visualization of clinical data.

Acknowledgements

We gratefully acknowledge financial support for this work from the Chalmers Foundation and the Knut and Alice Wallenberg Foundation. We thank Sergio Bordel, Rahul Kumar, Anastasia Krivoruchko and António Roldão for valuable discussions and editorial comments.

Conflict of interest statement

The authors declare no conflicts of interest.

References

1 Nielsen J. Systems biology of lipid metabolism: from yeast to human. *FEBS Lett* 2009; **583:** 3905–13.
2 Bruggeman FJ, Westerhoff HV. The nature of systems biology. *Trends Microbiol* 2007; **15:** 45–50.
3 Stephanopoulos G, Alper H, Moxley J. Exploiting biological complexity for strain improvement through systems biology. *Nat Biotechnol* 2004; **22:** 1261–7.

4 Choi S, ed. *Introduction to Systems Biology*. Totowa, NJ: Humana Press Inc, 2007.
5 Petranovic D, Nielsen J. Can yeast systems biology contribute to the understanding of human disease? *Trends Biotechnol* 2008; **26:** 584–90.
6 Petranovic D, Tyo K, Vemuri GN, Nielsen J. Prospects of yeast systems biology for human health: integrating lipid, protein and energy metabolism. *FEMS Yeast Res* 2010; **10:** 1046–59.
7 Hawkins RD, Hon GC, Ren B. Next-generation genomics: an integrative approach. *Nat Rev Genet* 2010; **11:** 476–86.
8 Ahn AC, Tewari M, Poon CS, Phillips RS. The clinical applications of a systems approach. *PLoS Med* 2006; **3:** e209.
9 Patil KR, Nielsen J. Uncovering transcriptional regulation of metabolism by using metabolic network topology. *Proc Natl Acad Sci USA* 2005; **102:** 2685–9.
10 Lewis TG. *Network Science: Theory and Practice*. Hoboken, NJ: John Wiley & Sons, 2009.
11 Gehlenborg N, O'Donoghue SI, Baliga NS *et al.* Visualization of omics data for systems biology. *Nat Methods* 2010; **7:** S56–68.
12 Auffray C, Chen Z, Hood L. Systems medicine: the future of medical genomics and healthcare. *Genome Med* 2009; **1:** 2.
13 Popel AS, Hunter PJ. Systems biology and physiome projects. *Wiley Interdiscip Rev Syst Biol Med* 2009; **1:** 153–8.
14 Kell DB. Systems biology, metabolic modelling and metabolomics in drug discovery and development. *Drug Discov Today* 2006; **11:** 1085–92.
15 Mo ML, Palsson BO. Understanding human metabolic physiology: a genome-to-systems approach. *Trends Biotechnol* 2009; **27:** 37–44.
16 Sreekumar A, Poisson LM, Rajendiran TM *et al.* Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature* 2009; **457:** 910–4.
17 Yildirim MA, Goh KI, Cusick ME, Barabasi AL, Vidal M. Drug-target network. *Nat Biotechnol* 2007; **25:** 1119–26.
18 Ma HW, Goryanin I. Human metabolic network reconstruction and its impact on drug discovery and development. *Drug Discov Today* 2008; **13:** 402–8.
19 Nicholson JK. Global systems biology, personalized medicine and molecular epidemiology. *Mol Syst Biol* 2006; **2:** 52.
20 Deisboeck TS. Personalizing medicine: a systems biology perspective. *Mol Syst Biol* 2009; **5:** 249.
21 Campion J, Milagro FI, Martinez JA. Individuality and epigenetics in obesity. *Obes Rev* 2009; **10:** 383–92.
22 German JB, Roberts MA, Watkins SM. Genomics and metabolomics as markers for the interaction of diet and health: lessons from lipids. *J Nutr* 2003; **133:** 2078s–83s.
23 Haslam DW, James WPT. Obesity. *Lancet* 2005; **366:** 1197–209.
24 Zimmet P, Thomas CR. Genotype, obesity and cardiovascular disease – has technical and social advancement outstripped evolution? *J Intern Med* 2003; **254:** 114–25.
25 Groop L, Orho-Melander M. The dysmetabolic syndrome. *J Intern Med* 2001; **250:** 105–20.
26 Cameron AJ, Magliano DJ, Zimmet PZ *et al.* The metabolic syndrome as a tool for predicting future diabetes: the AusDiab study. *J Intern Med* 2008; **264:** 177–86.
27 Loos RJF, Bouchard C. Obesity – is it a genetic disorder? *J Intern Med* 2003; **254:** 401–25.
28 Cobelli C, Man CD, Sparacino G, Magni L, De Nicolao G, Kovatchev BP. Diabetes: models, signals, and control. *IEEE Rev Biomed Eng* 2009; **2:** 54–96.
29 Pabinger S, Rader R, Agren R, Nielsen J, Trajanoski Z. MEMOSys: bioinformatics platform for genome-scale metabolic models. *BMC Syst Biol* 2011; **5:** 20.
30 Terzer M, Maynard ND, Covert MW, Stelling J. Genome-scale metabolic networks. *Wiley Interdiscip Rev Syst Biol Med* 2009; **1:** 285–97.
31 Akesson M, Forster J, Nielsen J. Integration of gene expression data into genome-scale metabolic models. *Metab Eng* 2004; **6:** 285–93.
32 Cakir T, Patil KR, Onsan Z, Ulgen KO, Kirdar B, Nielsen J. Integration of metabolome data with metabolic networks reveals reporter reactions. *Mol Syst Biol* 2006; **2:** 50.
33 Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO. Integrating high-throughput and computational data elucidates bacterial networks. *Nature* 2004; **429:** 92–6.
34 Patil KR, Akesson M, Nielsen J. Use of genome-scale microbial models for metabolic engineering. *Curr Opin Biotechnol* 2004; **15:** 64–9.
35 Gianchandani EP, Papin JA, Price ND, Joyce AR, Palsson BO. Matrix formalism to describe functional states of transcriptional regulatory systems. *PLoS Comput Biol* 2006; **2:** e101.
36 Feist AM, Herrgard MJ, Thiele I, Reed JL, Palsson BO. Reconstruction of biochemical networks in microorganisms. *Nat Rev Microbiol* 2009; **7:** 129–43.
37 Thiele I, Palsson BO. A protocol for generating a high-quality genome-scale metabolic reconstruction. *Nat Protoc* 2010; **5:** 93–121.
38 Palsson B. Metabolic systems biology. *FEBS Lett* 2009; **583:** 3900–4.
39 Bordel S, Agren R, Nielsen J. Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput Biol* 2010; **6:** e1000859.
40 Nookaew I, Gabrielsson BG, Holmang A, Sandberg AS, Nielsen J. Identifying molecular effects of diet through systems biology: influence of herring diet on sterol metabolism and protein turnover in mice. *PLoS ONE* 2010; **5:** e12361.
41 Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* 2010; **38:** D355–60.
42 Romero P, Wagg J, Green ML, Kaiser D, Krummenacker M, Karp PD. Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol* 2005; **6:** R2.
43 Croft D, O'Kelly G, Wu GM *et al.* Reactome: a database of reactions, pathways and biological processes. *Nucleic Acids Res* 2011; **39:** D691–7.
44 Oprea TI, May EE, Leitao A, Tropsha A. Computational systems chemical biology. *Methods Mol Biol* 2011; **672:** 459–88.
45 EBI. Rhea manually annotated database of chemical reactions.: Swiss-Prot Group at the Swiss Institute of Bioinformatics (SIB) and the Panda Chemoinformatics and Metabolism Group at the EBI. 2011.
46 Barrett T, Troup DB, Wilhite SE *et al.* NCBI GEO: archive for functional genomics data sets-10 years on. *Nucleic Acids Res* 2011; **39:** D1005–10.
47 Zheng-Bradley X, Rung J, Parkinson H, Brazma A. Large scale comparison of global gene expression patterns in human and mouse. *Genome Biol* 2010; **11:** R124.
48 Uhlen M, Oksvold P, Fagerberg L *et al.* Towards a knowledge-based human protein atlas. *Nat Biotechnol* 2010; **28:** 1248–50.
49 Wishart DS, Knox C, Guo AC *et al.* HMDB: a knowledgebase for the human metabolome. *Nucleic Acids Res* 2009; **37:** D603–10.
50 Flicek P, Amode MR, Barrell D *et al.* Ensembl 2011. *Nucleic Acids Res* 2011; **39:** D800–6.

51 Maglott D, Ostell J, Pruitt KD, Tatusova T. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res* 2011; **39:** D52–7.
52 Apweiler R, Martin MJ, O'Donovan C *et al.* Ongoing and future developments at the universal protein resource. *Nucleic Acids Res* 2011; **39:** D214–9.
53 Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M. The KEGG resource for deciphering the genome. *Nucleic Acids Res* 2004; **32:** D277–80.
54 Kanehisa M, Goto S, Hattori M *et al.* From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* 2006; **34:** D354–7.
55 Uhlen M, Bjorling E, Agaton C *et al.* A human protein atlas for normal and cancer tissues based on antibody proteomics. *Mol Cell Proteomics* 2005; **4:** 1920–32.
56 Ponten F, Jirstrom K, Uhlen M. The human protein atlas – a tool for pathology. *J Pathol* 2008; **216:** 387–93.
57 Ponten F, Gry M, Fagerberg L *et al.* A global view of protein expression in human cells, tissues, and organs. *Mol Syst Biol* 2009; **5:** 337.
58 Lundberg E, Uhlen M. Creation of an antibody-based subcellular protein atlas. *Proteomics* 2010; **10:** 3984–96.
59 Duarte NC, Becker SA, Jamshidi N *et al.* Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci USA* 2007; **104:** 1777–82.
60 Ma H, Sorokin A, Mazein A *et al.* The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol Syst Biol* 2007; **3:** 135.
61 Hao T, Ma HW, Zhao XM, Goryanin I. Compartmentalization of the Edinburgh human metabolic network. *BMC Bioinformatics* 2010; **11:** 393.
62 Shlomi T, Cabili MN, Herrgard MJ, Palsson BO, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol* 2008; **26:** 1003–10.
63 Vo TD, Greenberg HJ, Palsson BO. Reconstruction and functional characterization of the human mitochondrial metabolic network based on proteomic and biochemical data. *J Biol Chem* 2004; **279:** 39532–40.
64 Gille C, Bolling C, Hoppe A *et al.* HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol* 2010; **6:** 441.
65 Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol* 2010; **6:** 401.
66 Lewis NE, Schramm G, Bordbar A *et al.* Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol* 2010; **28:** 1279–85.
67 Bordbar A, Lewis NE, Schellenberger J, Palsson BO, Jamshidi N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol* 2010; **6:** 422.
68 Karlsson FH, Nookaew I, Petranovic D, Nielsen J. Prospects for systems biology and modeling of the gut microbiome. *Trends Biotechnol* 2011; **29:** 251–8.
69 Karlsson FH, Ussery DW, Nielsen J, Nookaew I. A closer look at bacteroides: phylogenetic relationship and genomic implications of a life in the human gut. *Microb Ecol* 2011; **61:** 473–85.
70 Saxena A, Carninci P. Whole transcriptome analysis: what are we still missing? *Wiley Interdiscip Rev Syst Biol Med* 2010; **3:** 527–43.
71 Reed JL, Patel TR, Chen KH *et al.* Systems approach to refining genome annotation. *Proc Natl Acad Sci USA* 2006; **103:** 17480–4.
72 Hatzimanikatis V, Li CH, Ionita JA, Henry CS, Jankowski MD, Broadbelt LJ. Exploring the diversity of complex metabolic networks. *Bioinformatics* 2005; **21:** 1603–9.
73 Kumar VS, Maranas CD. GrowMatch: an automated method for reconciling in silico/in vivo growth predictions. *PLoS Comput Biol* 2009; **5:** e1000308.
74 Orth JD, Palsson BO. Systematizing the generation of missing metabolic knowledge. *Biotechnol Bioeng* 2010; **107:** 403–12.
75 Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW. GenBank. *Nucleic Acids Res* 2011; **39:** D32–7.
76 Pruitt KD, Tatusova T, Klimke W, Maglott DR. NCBI Reference sequences: current status, policy and new initiatives. *Nucleic Acids Res* 2009; **37:** D32–6.
77 Cochrane G, Bates K, Apweiler R *et al.* Evidence standards in experimental and inferential INSDC Third Party Annotation data. *OMICS* 2006; **10:** 105–13.
78 Sayers EW, Barrett T, Benson DA *et al.* Database resources of the national center for biotechnology information. *Nucleic Acids Res* 2011; **39:** D38–51.
79 Berman H, Henrick K, Nakamura H. Announcing the worldwide Protein Data Bank. *Nat Struct Biol* 2003; **10:** 980.
80 Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. ExPASy: the proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res* 2003; **31:** 3784–8.
81 Schneider M, Tognolli M, Bairoch A. The Swiss-Prot protein knowledgebase and ExPASy: providing the plant community with high quality proteomic data and tools. *Plant Physiol Biochem* 2004; **42:** 1013–21.
82 Seal RL, Gordon SM, Lush MJ, Wright MW, Bruford EA. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res* 2011; **39:** D514–9.
83 Amberger J, Bocchini C, Hamosh A. A new face and new challenges for online mendelian inheritance in man (OMIM (R)). *Hum Mutat* 2011; **32:** 564–7.
84 International Union of Biochemistry and Molecular Biology. Nomenclature Committee., Webb EC. *Enzyme nomenclature 1992: recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes.* San Diego; London: Academic Press. 1992.
85 McDonald AG, Boyce S, Tipton KF. ExplorEnz: the primary source of the IUBMB enzyme list. *Nucleic Acids Res* 2009; **37:** D593–7.
86 Scheer M, Grote A, Chang A *et al.* BRENDA, the enzyme information system in 2011. *Nucleic Acids Res* 2011; **39:** D670–6.
87 Wang Y, Xiao J, Suzek TO, Zhang J, Wang J, Bryant SH. PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res* 2009; **37:** W623–33.
88 Degtyarenko K, de Matos P, Ennis M *et al.* ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Res* 2008; **36:** D344–50.
89 Harkewicz R, Dennis EA. Applications of mass spectrometry to lipids and membranes. *Annu Rev Biochem* 2010; **80:** 301–25.

Correspondence: Jens Nielsen, Department of Chemical and Biological Engineering, Kemivägen 10, Chalmers University of Technology, SE 412 96, Gothenburg, Sweden.
(fax: +46(0)31 772 3801; e-mail: nielsenj@chalmers.se).